DOCUMENT RESUME

ED 302 566 TM 012 568

AUTHOR Henfield, Paul; Balmaceda, Madeline
TITLE Teacher Licensing: Local versus National Approaches.
PUB DATE Apr 88
NOTE 15p.; Paper presented at the Annual Meeting of the New England Educational Research Organization (Rockport, ME, April 1988).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Descriptive (141)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Elementary Secondary Education; *Licensing Examinations (Professions); *National Programs; Occupational Tests; *Teacher Certification; *Test Construction; Testing Programs; Test Items
IDENTIFIERS National Teacher Examinations; New York City Board of Examiners; Pre Professional Skills Tests

ABSTRACT

The comprehensive testing program used by the New York City Board of Examiners to develop tests for licensure of teachers is compared with national approaches to licensing. The first step in developing a licensing examination in New York is thorough job analysis leading to the preparation of task statements. These are used to develop lists of knowledge, skills, and abilities for the job. The blueprint for the examination is developed from these lists. Written tests are prepared and validated by subject matter experts. Interview tests are also developed and performance tests are prepared for some areas. These results are evaluated on an item-by-item basis to identify problems with the test. On the national level, paper-and-pencil tests are the most prevalent approach, with the Pre-Professional Skills Test (PPST) and the NTE Core Battery the most commonly used. Misuse of NTE scores has been the subject of much recent discussion. It is recommended that a written test not be the sole assessment instrument for teacher selection. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED302566

TEACHER LICENSING: LOCAL VERSUS NATIONAL APPROACHES

Paul Henfield Madeline Balmaceda

NEW YORK CITY BOARD OF EDUCATION

BOARD OF EXAMINERS

65 Court Street

Brooklyn, New York 11201

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

PAUL M. HENFIELD

MADELINE BALMACEDA

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Presented at the annual meeting of the New England Educational Research Organization, Rockport, Maine, April, 1988.

TM012568

A growing trend in improving the quality of education in the nation's schools is the reliance upon a testing program for the certification and licensure of teachers. As of April 1987, at least 44 states have implemented or are in the process of implementing a testing program to assess the capabilities of teachers. This practice is continuing to spread in the face of tremendous controversy surrounding this issue. Most professional organizations support the concept of some kind of evaluation process for prospective teachers before they are granted licenses to teach. According to a 1986 Gallup Poll, the public supports this position as well. Results of this poll revealed that 85% of respondents felt that individuals should be required to pass a test before they are allowed to teach. However, disagreement does exist -- disagreement which centers primarily on who should be tested and what methods should be used for assessment.

For many years, it was argued that it was unnecessary to test people for teaching positions, since they already had been graduated from a college. This argument has lost favor in recent years. As many school administrators can attest, possession of a college diploma does not necessarily confer the ability to teach. Also, more and more of the brighter students are being drawn away from education toward careers offering greater financial rewards. Furthermore, almost every other profession requires individuals to pass a test after completing their studies. For example, lawyers must pass a bar exam and physicians must pass board-certification exams. If teachers are to be considered professionals, it would follow that teachers should be expected to do the same.

Since the concern of this paper is certification and licensure, the focus will be on methods that can be used to evaluate prospective rather than veteran teachers. The issue of testing veteran teachers is a highly sensitive one that is basically political in nature. Morever, some educational organizations which oppose the testing of veteran teachers, the National Education Association for example, support the testing of prospective teachers. In fact, there is general agreement that prospective teachers should be subjected to some type of evaluation process before being allowed to teach.

Prior to discussing the problems associated with testing teachers, a brief clarification of the terms "licensure" and "certification" is in order. These terms are often used interchangeably; in fact, there is a clear distinction between the two. Governmental agencies "license" people who have demonstrated an ability to fulfill the responsibilities of a job adequately. Professional organizations "certify" people who meet the standards of their profession. This distinction becomes blurred when states grant licenses to individuals to teach, although the methods used to license are more appropriate for certification purposes.

The increased role being given to tests in the licensure and certification of teachers raises many questions about the validity and effectiveness of the methods that are currently in use. In August of 1986, a study conducted by the United States Office of Educational Research and Improvement (OERI) concluded that "few states have effective teacher testing programs." The findings suggested that only the "grossly

incompetent" were identified; however, there is little evidence to suggest that other unqualified candidates fail these tests.

In response to the resurgence of testing teachers and to the growing concern about the effectiveness and objectivity of such testing, the Carnegie Corporation has sponsored the Carnegie Forum on Education and the Economy to establish a National Board responsible for developing appropriate standards in this field. It is envisioned that this board, national in scope, would certify teachers based upon individuals' ability to instruct students. This would signal a departure from what is generally taking place in many jurisdictions' testing programs which rely solely on paper-and-pencil tests that measure basic skills. The Carnegie Forum's approach, geared to the assessment of one's ability to teach, would necessitate evaluation techniques which may include, but also transcend, paper-and-pencil instruments. This new approach would dictate an emphasis on such methods as simulations, interviews, performance tests or tailored tests to achieve desired results. General knowledge tests, such as the Core Battery of the NTE and the Pre-Professional Skills Tests (PPST), do not measure, to any significant degree, the specific subject matter that individuals will be required to teach. Even the Speciality Area Tests, which <u>do</u> assess subject matter knowledge, do not evaluate an individual's ability to teach or interact with students. Nor are they able to evaluate oral communication and interpersonal relations skills, skills which are prerequisites for being a good teacher. At present,

the New York City Board of Examiners is unique in that it conducts a comprehensive testing program that includes multiple types of test components in the licensure of teachers.

I. The Testing Process Employed by the NYC Board of Examiners

The examinations administered by the New York City Board of Examiners are developed in accordance with the Federal Uniform Guidelines on Employee Selection Procedures and the American Psychological Association's Standards for Educational and Psychological Testing. These guidelines dictate that all examinations be job-related and that all test development procedures be validated.

A. JOB ANALYSIS

The first step in the development of any licensing examination at the Board of Examiners is to conduct a thorough job analysis of the position. This involves collecting data concerning the tasks performed by incumbents. Data collection is performed in many ways. The decision concerning the methods of data collection to be used is based on several factors including the number of incumbents in the position, the level of the position (e.g., teacher vs. supervisory) and practical factors such as available time and funds. A job analysis begins with a review of background materials such as prior job analyses, job descriptions, curriculum bulletins and Chancellor's Regulations. Next, observations of incumbents are conducted. The incumbents to be observed are selected so that a cross-section of the position within the city is represented.

When the position of concern contains too few incumbents, observations are conducted of similar positions, and other methods of collecting job analysis data are employed.

In general, the observations are a descriptive record of the incumbent's task behavior. In addition, a record is kept of the amount of time spent on each task behavior as well as notes concerning materials used and persons with whom the observee interacts during the course of the observation.

In addition to observations, other methods of collecting job analysis data include interviews, questionnaires and logs. Participants for these methods are selected in the same manner as for observations in order to ensure representation of a cross-section of the population. Interviews and questionnaires are adminstered to individuals as well as groups of incumbents. Supervisors and other individuals related to the position being analyzed may be included in these procedures. The goal of all these methods of collecting data is to obtain as complete a picture of the tasks of the position as possible.

Next, the information gathered is put into the form of a list of task statements of the position. A panel of Subject Matter Experts (SMEs) is convened to validate these task statements. This involves a review of each task statement for its content as well as its form. Morever, each task statement is assigned by each panel member a global rating of importance that takes into account the factors of time, difficulty and

consequences of error. These ratings yield a final list of ranked task statements of the position. This process identifies the critical job tasks from which the knowledges, skills and abilities required of the position are derived.

After the list of ranked task statements is obtained, a list of knowledges, skills and abilities (KSAs) required in order to perform the tasks of the job is developed. This requires a careful examination of each job task in order to derive its accompanying KSAs. Once a comprehensive list of KSAs is established, they are reviewed, revised and rated by a panel of subject matter experts in the same manner as for task statements. From these individual ratings, a ranked list of KSAs is obtained. In order to ensure the job-relatedness of every examination, it is necessary to link each phase of the examination development process to the originally identified job tasks of the position. Each ranked knowledge, skill and ability is linked to the specific task statements from which it was derived, and a written record of these linkages is maintained. In addition, a cut-off is established based on the obtained mean and standard deviation of the KSA ratings. In general, those KSAs which have ratings greater than one standard deviation below the mean are not considered critical to the job and, therefore, are not tested for. These KSAs often include ones that are not required for entry-level applicants and that can be learned on the job.

B. EXAMINATION DEVELOPMENT

Based on the obtained ranked KSAs, the blueprint for the examination is then developed. Each critical KSA is evaluated as to how it can best be

assessed (e.g., written, interview, performance tests etc.) In this way, each KSA is assigned to a particular examination part or parts. Weights for the various parts of the examination are based on the number and importance of the KSAs assessed by each part. Adjustments are made for overlapping KSAs, that is, KSAs that will be measured by more than one part of the examination. The obtained weights are also reviewed by subject matter experts who offer their expertise concerning the appropriateness of the obtained weights.

1. Written Tests

The specific content of each written examination is determined by the assignment of specific KSAs to each examination part. After the critical KSAs for the written test have been identified, subject matter experts, trained in item-writing principles and techniques, construct the items. In addition, relevant items that have been used on previous secure examinations are retrieved from item banks if their performance on previous examinations have been found to be reliable. This creates a pool of items from which the final test items will be selected. After each item is carefully edited by in-house staff, they are validated by a panel of subject matter experts. This involves a review and revision of each item and a rating of each item in terms of its job-relatedness and importance. Based on the results of these ratings, items are selected for the final version of the test. Special care is taken to ensure that the proportion of items chosen for the test reflect the number and importance of the KSAs on the established test blueprint.

Passmarks for the test are determined by a modified Angoff procedure in which a panel of subject matter experts is asked to indicate for each test item the percentage of entry-level applicants they believe will correctly answer each item. In addition, after rating each individual item, each panel member is asked to assign a global percentage indicating the percent of items on the total test that will be answered correctly by entry-level applicants. The means of the individual and global ratings are compared, and a passmark is then determined, rounded to the nearest 5 point interval.

After multiple-choice questions are scored, an item analysis is obtained for tests involving thirty or more applicants. A comprehensive review of item analysis data is conducted. Any problematic items (e.g., negative point-biserial correlations) are reviewed and, if appropriate, deleted. The test is then rescored.

For essays and in-basket questions, rating guides, or standard answers, are prepared. Those subject matter experts who will prepare the rating guides are trained in the techniques involved in writing model responses. After the responses are constructed, the rating guides are validated by a panel of subject matter experts in terms of their completeness and correctness. Credit allocation is also assigned to each model response. Before the actual rating of applicant responses takes place, interrater reliability is determined to ensure comparability of ratings among raters.

2. Written English Examination

The Board of Examiners, for every license examination, administers a Written English test. Applicants are asked to write an essay in continuous discourse about a given topic. These examinations are graded for grammar, spelling and language usage only. A demerit scoring approach is used, as opposed to a holistic approach, because the reliability of the latter has not been confirmed.

3. Interview Tests

As with all other types of examinations, the interview test measures those knowledges, skills and abilities that have been assigned to this examination part of the test blueprint. In the interview test situation, in the presence of an examining panel, applicants are required to respond to a set of questions designed to measure appropriate KSAs, and also to respond to related follow-up questions. Prior to the interview test, examination panel members are given an extensive briefing concerning the conduct of the test and appropriate evaluation and scoring procedures. In addition, interrater reliability is determined at the briefing session.

4. Performance Tests

Based on the job analyses of some license areas, a performance test is determined to be the most appropriate measure of certain critical KSAs. For example, in some of the trade areas or in areas requiring diagnosis and remediation of student difficulties, performance tests have proven to be an appropriate measure of an applicant's minimum competency. These

examinations can be administered in a school situation or in a simulated environment.

5. Post-test Procedures

As part of the test item review process, the Board of Examiners evaluates its test results on an item-by-item basis to identify possible problems with adverse impact. This entails an examination of item performance broken down by ethnicity and gender. For those items where a disparity in performance is found, a reevaluation by subject matter experts and psychometricians takes place. If problems relating to adverse impact are found, the item is deleted and the tests are rescored.

In addition, the Board of Examiners employs a comprehensive appeals process for all its examinations. Applicants are given the opportunity to challenge particular short-answer items as well as the performance and interview tests. Each appeal is reviewed by an independent evaluator who then makes a recommendation to the Board of Examiners, which in turn, rules on the appeal.

II. National Approaches

The most prevalent approach used by most states today in the licensing of teachers relies on standardized paper-and-pencil instruments. The most common instruments are the Pre-professional Skills Test (PPST) and the National Teachers Examinations (NTE), both produced by the Educational Testing Service (ETS). The PPTS is a general skills test which assesses

basic reading, writing and mathematics. It is used in approximately six states and takes 2 1/2 hours to complete.

The NTE is the most frequently used test for teachers and is composed of two elements: The Core Battery and approximately 27 Specialty Area tests. Some part of the NTE battery is used in approximately 22 states. The Core Battery consists of 340 multiple-choice items and one essay item which form three separate tests: 1. Communication Skills, which assesses reading, listening and writing skills 2. General Knowledge, which includes items on literature, arts, mathematics, science and the social sciences 3. Professional Knowledge which assesses pedagogy and student evaluation.

There are 27 Specialty Area exams with new exams being prepared on an ongoing basis. These exams are available in most of the major subject areas, for example, English, Math, Music, Special Education, Speech and Foreign Language.

The NTE has encountered severe criticism in recent years as a valid instrument for the licensing of teachers. There have been legal suits in several states challenging the use of the NTE as a teacher certification instrument because it has evidenced adverse impact against minority candidates. ETS itself has stated that it would be inappropriate to use the NTE as the sole criterion in the licensure of teachers. Further,

they state that the "use of cutoff scores ... is generally discouraged and that use of such scores in determining a teacher's retention, tenure or salary level is a misuse of the examination."

Until recently, New York State education law mandated that certain teaching vacancies in New York City could be filled via an alternate route to the regular Board of Examiners' testing process. This shorter process, implemented to prevent teacher shortages in the lower-performing schools, enabled individuals to receive teaching licenses based solely on NTE scores.

In March, 1988 ETS announced that it would block the New York City Board of Education from issuing licenses through the alternate route. The service said that New York City was misusing its test because cutoff scores were determined by improper procedures. New York City is not unique in this respect. Results of an ETS survey indicate that approximately 70% of the institutions that receive NTE scores from the service use these scores improperly. However, most of these institutions used the NTE to confirm state certification and not to license teachers.

These findings suggest the following question: How should states and school systems using a standardized test for teacher certification proceed? Those school systems that have the resources can either develop their own testing programs tailored to their specific needs or commission a testing service to develop one for them. Lastly, additional selection

criteria should be incorporated into the process, for example, an interview or performance test, since it is not recommended that a written test be used as the sole assessment instrument governing a teacher selection program.